

## Transparency and Building Lexical Dependency Graphs

CHARLES J. FILLMORE<sup>i</sup> and HIROAKI SATO<sup>ii</sup>

<sup>i</sup>International Computer Science Institute, Berkeley and <sup>ii</sup>Senshu University, Japan

### 0. Introduction

The authors are participants in a project, called “FrameNet”<sup>1</sup>, which is aimed at building a large computer lexicon of contemporary written English and making it accessible through the World Wide Web for both computational and lexicographic interests. In the process of designing this resource, we have to keep in mind how it can serve its intended applications. A prerequisite to most imaginable NLP applications is *word sense disambiguation* (WSD), the automatic process by which a word in a linguistic context can be (probabilistically) assigned its locally intended meaning. This paper will introduce FrameNet and will characterize a facility, based on its tools and data, that could in principle be directed to WSD efforts, and will suggest how both technical (engineering) and linguistic considerations must be called on to build it.

Briefly (a more detailed account can be found below), the manner in which the FrameNet work is carried out, one word at a time, involves the annotation of sentences exemplifying each word taken from a large corpus<sup>2</sup>, where each resulting annotation provides an example of the use and the essential combinatorial properties of one *lexical unit*<sup>3</sup> (LU). The overall goal is to offer for each LU a representative collection of annotated examples upon which automatic processes can operate to display a *valence description*<sup>4</sup> of this particular word, in

<sup>1</sup> FrameNet is an NSF-sponsored resource-building effort supported under Grant No. ITR/HCI – 96132: “FrameNet ++”: An On-Line Lexical Semantic Resource and Its Application to Speech and Language Technology. PIs are C Fillmore and S Narayanan (ICSI), D Jurafsky (U Colorado), and M Gawron (San Diego State U). Project Manager is C Baker (ICSI). Current funding continues through August 2003. Information at <http://framenet/icsi/berkeley/edu/~framenet>.

<sup>2</sup> The current FrameNet Corpus is the British National Corpus, which we are using courtesy of Oxford University Press and with software tools made available through the Institut für Maschinelle Sprachverarbeitung of the University of Stuttgart. Information about the BNC can be found at <http://info.ox.ac.uk/bnc/>

<sup>3</sup> A lexical unit is a lexeme in one of its senses; we follow the usage of D. A. Cruse (1986).

<sup>4</sup> A FrameNet valence description is a display of the “frame elements” (FEs, roughly participant roles) associated with a lexical unit in a given frame, and the manner of their syntactic realization,

terms of the semantic roles, grammatical functions and phrase types of those of the word's phrasal companions that fill slots associated with the word's semantic frame.

Manual annotation of large corpora in the FrameNet manner would be prohibitively expensive, and it is not efficient to let the number of annotations reflect the relative frequency of words, word senses, or valence patterns, and so attempts are being made to automate the process in a way that will allow at least some of the tagging to be done on a large scale using both machine learning and human-written apriori rules. The ability to conduct automatic semantic/syntactic annotation would be enhanced by the study of lexical collocations within the clusters of information that represent the FrameNet annotation of individual LUs (because it could assist in disambiguation of polysemous words), and at the same time the automatic annotation of new documents could help us discover new collocations. The automated ability to find such collocations could make it possible: 1) to recognize collocations for their own sake, accumulating collocation frequencies across a wide variety of corpora<sup>5</sup>; and 2) to detect the specific lexical collocations that occur in a given document, which could be a clue to the topic of the individual passages in the document. The FrameNet database as presently constructed aims at coverage of the distributional varieties, but cannot directly yield information about relative frequencies. If such software can be created, we hope to compile frequency evidence for lexical collocations and/or semantic type selection for each predicating LU.<sup>6</sup>

The collocations will be collected in the form of what can be called *kernel dependency graphs* (KDGs)<sup>7</sup>. These are small packages of information that associate the lexical head (governor) of a set of related dependents, the lexical heads of the constituents that are dependent on that governor, and the frame-specific semantic relations by which the dependent elements are related to the governor. A large corpus-based registry of KDGs would permit queries about, say, the most common dependents of given governors (in given semantic roles), the most common governors of particular nouns, and so on.

---

this expressed in terms of grammatical functions and phrase types if they are present in the sentence, and type of omission (depending on what licenses their absence in the sentence) if they are not realized in the sentence.

<sup>5</sup> The registry of lexical collocations that FrameNet is able to accumulate will be a skimpy but semantically and syntactically richer variety of the Dependency Resource of Dekang Lin (<http://www.cs.ualberta.ca/~lindek/demos.htm>). The differences are that the collocates will be chosen according to semantic role, rather than merely syntactic relations, and they will differ across different senses of the target word. (Further differences will be described in the present paper.)

<sup>6</sup> We are only minimally able to carry out such studies at the present time; our current efforts are to create the needed conceptual and software infrastructure.

<sup>7</sup> For present purposes the assumed grammatical model is of a Dependency type (Tesnière 1959), but the term "kernel" is borrowed from early transformationalist writing (Chomsky 1957) referring to the structure of a predication with a single verb and its accompanying grammatical partners. The actual characteristics of our KDGs differ from both, precisely because of the treatment of the transparency structures among other things.

The purpose of the present paper is to argue that for such KDGs to be semantically relevant, we need to be able to recognize certain kinds of discrepancies between syntactic and semantic structure, and this involves recognizing the role of 1) *support verbs* in the case of nominal governors—nouns that have their own frame structure—and 2) various classes of *transparent nouns*<sup>8</sup>, so that we can spot instances of patterns such as [Noun+of+Noun] in which it is the semantic head (here, the second noun), not the syntactic head (the first noun) that is most related to the context of the NP as a whole in respect to collocations, selection or agreement phenomena.

### 1. Acquisition of an Inventory of Kernel Dependency Graphs

The objects we intend to accumulate in the KDG project are clusters of lexical items, each serving as a lexical head<sup>9</sup>, seen in a quasi-dependency grammar representation. One such cluster could be a simple transitive verb in association with the head nouns of the constituents serving as its subject and its direct object. However, we wish to give more information than can be found in such a minimally structured cluster of words. We also wish to identify 1) the semantic frame within which the governor serves as an LU (the frame that provides the elements of its conceptual structure), 2) markers of the syntactically oblique arguments in the cluster, and 3) the semantic roles (called *frame elements*) of each dependent. The purpose of the present paper is to lay out the grammatical and technical issues in the effort to achieve this.

Each of the clusters of information that we hope to assemble as the work of the KDG project will contain 1) a lexical head, i.e., a predicator of some kind, an LU that evokes a semantic frame with one or more arguments<sup>10</sup>, and 2) the lexical heads of the constituents that express the arguments of the head predicator. For example, one such cluster (reduced to the lexical heads), based on the verb *steal*, would be derived, as the display in (2), from sentence (1). (The head word is the governor, the indented words are the lexical heads of the dependents.)

(1) The boy we noticed entering the store behind us was caught stealing a fish.

(2) steal  
    boy  
    fish

From the same sentence, another such cluster would be that seen in (3).

<sup>8</sup> The term is said to be due to Naomi Sager, but we have not found the source.

<sup>9</sup> In general, established compounds will be treated as single lexical heads. Named entities, such as personal names, place names, dates, etc., will be converted to names of types in the collocation registry.

<sup>10</sup> Since the information we seek involves constituents that fill out some of the details of the frame of the governor, we needn't begin by establishing criteria for an argument vs. adjunct distinction.

- (3) catch  
       SOMEONE  
       boy  
       stealing

The displays in (2) and (3) are incomplete. Versions that include frame identity, and the marking and semantic role information, can be seen in (4) and (5). The verb *steal* belongs to, or evokes, a Theft frame, and the elements of that frame include the *perpetrator* and the *loot*.<sup>11</sup> The diagram<sup>12</sup> identifies the governor, the frame, and those participants of the frame realized in the sentence.

- (4) <KDG rdf:ID="1137864">  
       <governor>steal</governor>  
       <frame rdf:resource="Theft">  
           <perpetrator>boy</perpetrator>  
           <loot>fish</loot>  
       </frame>  
     </KDG>

The governor *caught* in that same sentence evokes what we might call a Spotting frame, which requires an understanding of an observer, the person observed, and the act whose performance has been observed.

- (5) <KDG rdf:ID="46823">  
       <governor>catch</governor>  
       <frame rdf:resource="Spotting2">  
           <observer>SOMEONE</observer>  
           <observed>boy</observed>  
           <act>stealing</act>  
       </frame>  
     </KDG>

It can be noticed (1) that *boy* was taken as the "perpetrator" of the *steal* governor, though this is through indirect syntactic evidence (the NP headed by *boy* is the direct object of *caught*, but construed through familiar control structures, as satisfying the subject role of *caught*'s complement VP), and (2) that the agent of the *catch* governor is missing, but understood as an indefinite someone, as part of the interpretation of the passive voice in *was caught*.

Thus we see, a part of the task of deriving KDG clusters from a corpus involves getting around the lexico-syntactic structures that can conceal the "deep" syntactic arrangements within the predication or that can intervene between a governor and its arguments.

<sup>11</sup> For present purposes we assign no special significance to the names of frames or the names of the frame elements. It is only important that the frame names are unique and carefully defined, and that the frame element names are defined relative to the frames to which they belong.

<sup>12</sup> Since the FrameNet data will be presented to the world in both XML and DAML+OIL framework, the display here is only to give an idea what the formal representation will look like. The ID numbering and the resource names suggested here are not actually a part of our present database.

Something approaching a complete semantic analysis of each sentence could be built by carrying this through with each predicator in each clause, showing all possible semantic links of predication, modification, coordination, etc. (The "act" in KDG (5) is headed by the *steal* of KDG (4), so a representation of the whole complex, at this level, could be built by binding arguments of one KDG with those of another, or by binding one argument with another entire KDG.) But here we are defining a narrower task, allowing ourselves to ignore certain syntactically higher predicates and find the "deepest" dependency graphs in a clause, in the hope that this would make it possible to zero in on a semantic structure that could reliably indicate something about the topic of the passage. The intuition is that if we are interested in detecting the core idea in a passage, from a sentence like *Authorities have revealed that witnesses claimed that an accountant embezzled an entire month's payroll from Procter and Gamble*, we would be more interested in claims about what somebody has *stolen* than in claims of what some people have *said*.

For verb-headed KDGs, the identification of controlled arguments is fairly straightforward, given familiar accounts of syntactic control. If FrameNet annotations were accompanied by complete successful parses of the sentences, it would not be necessary for FrameNet annotators to identify those constituents external to the VP which realize FEs of the head-verb's frame. But for the sake of being able to collect collocational and selectional information, i.e., for the sake of being able to answer the questions, which words [collocations] and which classes of words [selectional preferences] occur in particular FE functions with given frames, we seek to label all constituents in the selected sentences which stand for, on the basis of obligatory syntactic relationships, the various FEs of each target verb. This means not only recognizing constituents which are "extracted" (e.g., through topicalization, interrogation, or relativization) but also the arguments of embedding predicates (verbs, nouns, or adjectives) which also serve as arguments the VPs of their complements. For the KDG study, in fact, an assignment that we could justify would have us ignore the control and raising predications themselves, except for the sake of finding KDG arguments, wherever the semantic contribution of the embedding predicates is limited to such matters as Appearance (*seem*), Probability and Ability (*likely, may, can, ability*), Intention (*try, want, eager, willing, intention, decision*), Report (*claim, say, announcement*), Belief and Expectation (*believe, expect*), Phase or Aspect (*begin, continue, stop*), Obligation (*must, ought*), Causation and Inducement (*cause, persuade*); in other words, almost all of the familiar control and raising verbs and adjectives, including the modals, plus complement-taking nouns indicating any of the listed categories.

### 1.1 Nouns as Frame Evokers

One way in which the KDG project goes beyond the simplest notion of kernel structure is that we are equally interested in the frame structure of nouns as well as verb. The simplest case of an argument structure centered in a noun is that of

nominalizations of verbs that provide those same structures. The FEs that go with *decision* are the same as those that go with *decide*, as suggested by the examples in (6).

- (6) a. we decided that we would storm the palace
- a'. the decision that we would storm the palace
- b. I decided to confess everything
- b'. my decision to confess everything

We see from the examples in (6) that clausal and phrasal direct complements of the noun are essentially those of the source verb. However, it is possible to realize the subject of the corresponding verb, standing for the one who makes the decision, as a postnominal oblique complement of the noun, i.e., with *of* or *by*, as in (7a) and (7b), or as a prehead noun modifier as in (7c).

- (7) a. the decision of the committee
- b. the decision by the committee
- c. the White House decision

A Topic FE can also be recognized with this noun, as postnominal oblique, as in (8a), or as a pronominal noun or adjectival modifier, as in (8b).

- (8) a. a decision concerning her retirement
- b. a retirement decision

But there is an additional device for identifying a participant in a noun-evoked frame: the *support verb*<sup>13</sup> (SV). Support verbs in the case of event nouns are analogous to the control verbs that provide VP-external arguments for nonfinite verbs, in that one of their arguments is construed as a participant in the situation associated with the head noun in their direct object. In the case of *decision*, the support verb is *make* (in U.S. English) or *take* (in U.K. English). In sentence (9), it is understood that the subject of *make* is the decision-maker in the situation identified by the word *decision*.

- (9) I *made* the premature decision that it was time to introduce myself to her father.

Obviously, not every verb that can take a nominalization as its object has this function. Those italicized in (10), for example, do not.

<sup>13</sup> The term appears to have been first used by the late Maurice Gross, but the concept has been around for a long time. An earlier German name was Funktionsverb. Our use of the term is broader than the traditional concept, including a number of verbs that have semantic content of their own.

- (10) a. We just have to *accept* your decision.  
b. They tried to *force* a decision.

The criterion for being a support verb is that of linguistic necessity the subject of the verb must be understood as a participant in the event designated by the supported noun.

The traditional SVs (especially those called *light verbs*) have little semantic responsibility beyond allowing the frame-structure of the noun to be expressed in a verb-headed clause: in this limited function we find in English mostly the verbs *make*, *have*, *give*, *take*, and *do*. But there are many other SVs that have semantic functions analogous to those of control verbs. Most, or perhaps all, of these, fit various categories of lexical functions elaborated by Igor' Mel'cuk (1996).

For some there is not much more to say than that the subject of the verb is the Actor of the noun's event type: *say a prayer*, *sing a song*, *tell a lie*, plus the light verbs. Different SVs can select different perspectives on events, such as the Actor versus Undergoer perspective seen in *give* vs. *take an examination*, *submit* vs. *receive an application*; *inflict* versus *sustain an injury*; *perform* versus *undergo an operation*. In some cases the SV selects one or another sense of a word: thus *have an argument* selects the 'quarrel' meaning of the noun; *make an argument* selects the 'reasoning' sense. Some SVs differ subtly in register, as in *make a complaint* versus *register*, *lodge* or *enter a complaint*. Some SVs go with nouns in particular semantic classes: names of sins and crimes, as well as the words *sin* and *crime*, seem to prefer *commit*. Some SVs require their subjects to be participants in the larger scenario than the event that is directly associated with the verbal source of a noun: thus *make a promise* is like the verb *promise*, but *keep a promise* and *break a promise* do not themselves refer to the act of promising; similarly, while *give an exam* and *take an exam* are related to the event expressed by the verb *examine*, *pass an exam* and *fail an exam* are not. But they are still bound to arguments of the event associated with the noun.

The KDG project, insofar as it aims to record KDGs built around noun-expressed frames, clearly needs access to information about SVs appropriate to each frame-evoking noun.

## 1.2 Transparent Nouns

The second departure from the simplest version of our task involves situations in which there is a discrepancy between the syntactic head of a phrase and its semantic core. In many structures of the type [N of N], the first noun has some quantifying or typing function and it is the second noun<sup>14</sup> that bears collocational relations with something in the surrounding text. Those instances of first-nouns in this pattern that have this property we will call *transparent nouns*. Their semantic

<sup>14</sup> A tighter formulation would acknowledge the possibility of recursion: in *I could never drink more than a pint of this kind of beer*, we have, with *pint* and *kind*, one [N of N] structure inside another, and it is the deepest one that serves as a collocate with *drink*.

functions include those of Quantities, Aggregates, Parts & Portions, Types, Classifiers, and Evaluators. The words that serve this function are not grammatically dedicated to serving in this way, and some of them can stand on their own with *of*-phrases as their complements. A contrast between the two situations can be seen in item (11) where the word *number* can be seen as having either function.

- (11) a. We tried to estimate the *number* of apples the horse ate.  
b. The horse appears to have eaten a *number* of poisoned apples.

In (11a) an important collocational relation holds among *we*, *estimate* and *number*; in (11b) among *horse*, *eat* and *apples*, the latter being transparent to the presence of the word *number*. In (11b) the word *number* is paired with *lot*, *bunch*, etc.; in (11a) its partners include *quantity*, *size*, *magnitude*, etc.

The intended idea of transparency can be seen by considering how one might derive a semantically relevant KDG from a sentence like (12):

- (12) A majority of tobacco producers use a form of asbestos in this kind of filter.

If we took the lexical heads of each of the NP constituents in this sentence, the KDG, using the words underlined in (13), would be (14)<sup>15</sup> (majority uses form in kind).

- (13) A majority of tobacco producers use a form of asbestos in this kind of filter.

- (14)
- ```
<KDG rdf:ID="11786424">
  <governor>use</governor>
  <frame rdf:resource="use3">
    <agent>majority</agent>
    <ingredient>form</ingredient>
    <product>in: kind</product>
  </frame>
</KDG>
```

By skipping the transparent nouns of quantity and type, choosing instead the nouns underlined in (15), we derive KDG (16), which we take consider more informative as to the point of the passage (tobacco-producers use asbestos in filter).

- (15) A majority of tobacco producers use a form of asbestos in this kind of filter.

<sup>15</sup> We take the frame associated with *use* in this sentence to involve an agent who consuming ingredients (resources, material) in creating from them some sort of product.



(16) 

```
<KDG rdf:ID="11786424">
  <governor>use</governor >
  <frame rdf:resource="use3">
    <agent>tobacco producers</agent >
    <ingredient>asbestos</ingredient>
    <product>in: filter</product>
  </frame>
</KDG>
```

Some examples of collocational relations obtaining through various kinds of transparent nouns are given in Table 1.

Table 1

Prep & favored object	on the shelf in the closet	on part of the shelf in part of the closet
Nouns & required possessors	her husband my wife	her idiot of a husband my gem of a wife
Subjects & favored VPs	water would quench my thirst	a bit of water would quench my thirst
Verb & favored object	hard to iron this shirt	hard to iron this kind of shirt
Adjective & favored complement	guilty of murder	guilty of three counts of murder
Noun & favored support verb	made a good impression	made that kind of impression

Notice that the sentence in (17), involving a noun derived from the verb *use*, would yield essentially the same structure as the KDG in (16), although this time the governor is a noun, and the agent is connected with the frame through a support verb those subject is controlled through an embedding predicate of saying.

(17) A number of tobacco producers are said to have made use of some form of asbestos in this kind of filter.

## 2. Introducing FrameNet

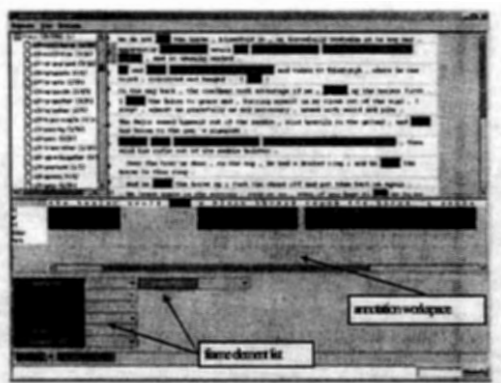
There are several aspects of the FrameNet data which are capable of serving the purposes of KDG derivation discussed in this paper.

The FrameNet project is producing a dictionary *cum* lexicon, for both human and computational uses, that documents the combinatorial properties of English lexical items, in semantic and syntactic terms, based on attestations in a very large corpus. Basically, the work of FrameNet is to discover and describe those semantic frames which provide the conceptual background for groups of semantically related words; to analyze the situational props, participants and components that conceptually make up part of the frame, to give them names (as "frame elements", FEs); to come up with a list of words—verbs, nouns,

adjectives—whose meanings evoke this frame; to explore attestations of some of these words in the Corpus to get a preliminary idea of the ways in which the various FEs combine with the word and how they are syntactically realized; to create a concordance of instances of each word in a way that is sure to include all varieties of the syntactic contexts in which the word occurs (in the intended meaning); and to submit these sentences to a software tool that can be used for selecting sentences that typify uses of the word (in the intended sense) and labeling the constituents that express or identify the FEs.<sup>16</sup>

Figure 1 gives a glimpse of the annotation tool used in the project. Shown is a set of sentences with the verb *tie* as analyzed in an Attachment frame. (The part of the sentence visible in the workspace is *the healer would tie a black thread round the horse's ankle*.)

Figure 1



The search for the support verbs that accompany the noun *complaint*, and the syntactic roles of its FEs, yields information like what is seen in Figure 2. (The inclusion of *afterwards* as a support verb is, of course, a mistake.) We can notice that the frame structure of this noun has separated Speaker as the person submitting the complaint and Addressee as the person toward whom the complaint is addressed, and the verbs *face* and *get* can have Addressee subjects, while *express* and *lodge* have Speaker as subject. (The verb *have* seems to allow both possibilities. *I have a complaint* may mean that I wish to express a complaint; *we have had numerous complaints about our prices* refers to complaints we have received.)

<sup>16</sup> In practice the process zigs and zags. Proposals are made for the structure of a new frame, and wordlists are compiled. Examination of concordance lines for some of the proposed words may require us to revise the frame description, or to add FEs, or to consider whether the initial list should be divided into more than one frame. After such decisions are made, some words get added to the lists and others get set aside.

Figure 2

The screenshot shows a web browser window titled "FramesQL search result". The browser's address bar and navigation buttons (Back, Forward, Stop, Refresh, Home, AutoFill, Print, Mail) are visible. The search results are displayed in a table with the following columns: No., frameIDs, frames, lemma, pos, goto2, and Concise Oxford Dic. 10.

No.	frameIDs	frames	lemma	pos	goto2	Concise Oxford Dic. 10
1	0*	Statement	complaint	N	entry	xxxx

Below the table, the following text is displayed:

FE KDG valence transparentN untra. N Supp Gov

Num sort = FE; complaint.

01 afterwards + complaint + send + complaint.N + Topic

01 lodge + complaint.N

01 make + Subject + complaint.N + Addressee

01 Message + Subject + complaint.N

01 Message + Subject + complaint.N + Topic

01 Addresser + face + complaint.N + Topic

01 Addresser + get + complaint.N + Subject

02 Addresser + have + complaint.N + Topic

01 Subject + express + complaint.N + Topic

01 Subject + have + complaint.N + Topic

01 Subject + lodge + complaint.N

Figure 3 shows two KDGs found in attestations of the verb *kidnap* displayed with the actual sentences on which they are based, and demonstrates that the quantity expression *number* was successfully ignored.

Figure 3

The screenshot shows a web browser window titled "FramesQL search result". The browser's address bar and navigation buttons (Back, Forward, Stop, Refresh, Home, AutoFill, Print, Mail) are visible. The search results are displayed in a table with the following columns: No., frameIDs, frames, lemma, pos, goto2, and Concise Oxford Dic. 10.

No.	frameIDs	frames	lemma	pos	goto2	Concise Oxford Dic. 10
1	0*	Statement	complaint	N	entry	xxxx

Below the table, the following text is displayed:

FE KDG valence transparentN untra. N Supp Gov

Num sort = KDG; kidnap, trans.

01 /businessmen + [kidnapped] + /for : ransom + [Place/in : manila] + [Perpetrator/CN]

01 /teachers + [kidnapped] + [Formulated/by : contrast]

2

01. [Top]: /businessmen + [kidnapped] + /for : ransom + [Place/in : manila] + [Perpetrator/CN]

[back to the main table]

1. 241808: [Victim] have been kidnapped<sup>PT</sup> and held for ransom [Place, in Manila] . [Perpetrator, CN] [Top]

01. [Top]: /teachers + [kidnapped] + [Formulated/by : contrast]

[back to the main table]

1. 241937: The dangers for teachers in war zones of being identified with the government were exemplified by the fate of [Victim] . [Perpetrator, CN] [Top]

Similarly, in Figure 4 the system has correctly identified *prints*, through the transparent noun *collection*, and *jackets* through *loads*.

Figure 4

FramesQL search result

No.	framesql	framesql	framesql	framesql	framesql	framesql	framesql	framesql	framesql
1	0	Theft	steal	V	entry	xxxx			

FE: KDG: valence: transparentN: untra. N (Supp: Gov)

Num	sort = KDG: Theft, steal, V, trans.	framesql
01	[Goods/prints] + stolen + [Perpetrator/by: thieves]	framesql
01	[Perpetrator/Hijackers] + stolen + [Goods/leather jackets]	framesql
2		framesql

01. [Top]: [Goods/prints] + stolen + [Perpetrator/by: thieves]

[back to the main table]

1. 196940: LGoods: A collection of prints depicting Whitby at the turn of the century have been stolen? [Perpetrator/by: thieves] . [Top]

01. [Top]: [Perpetrator/Hijackers] + stolen + [Goods/leather jackets]

[back to the main table]

1. 196815: LPerpetrator: Hijackers who had stolen LGoods: a load of leather jackets attracted attention by driving at 40mph on the M2 near Stungbourne, Kent . [Top]

### 3. Summary

The database-viewing tools demonstrated so far are not yet able to generate a full registry of KDGs from the FrameNet annotations, and are certainly not capable of extracting KDGs from raw text. But they do show that if a parser can recognize not only the familiar embedding structures for filling out the missing components of nonfinite VPs but also support verbs in our sense (which can be seen as having a similar function) and transparent nouns, the data needed for constructing a KDG registry are at hand.

### Acknowledgements

The authors are indebted to Srinu Narayanan and Collin Baker for help at various points in the preparation of this paper.

### References

- Akimoto, Minoji. 1989. *A Study of Verbo-Nominal Structures in English*. Shinozaki Shorin. Tokyo.
- Chomsky, A. Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Cruse, D. Alan. 1986. *Lexical Semantics*. Cambridge, U.K.: Cambridge University Press.
- Davis, Anthony R. & Leslie Barrett. 2001. Support verbs, nominalizations and participant roles. In Pierrette Bouillon & Kyoko Kanzaki (eds.) *Proceedings of the First International Workshop on Generative Approaches to the Lexicon*. Geneva.
- Fontenelle, Thierry. 1999. Semantic resources for word sense disambiguation: a sine qua non?. *Linguistica e Filologia*, 9: 25-41.

- Mel'cuk, Igor'. 1996. Lexical functions: a tool for the description of lexical relations in the lexicon. In Leo Wanner (ed.) *Lexical Functions in Lexicography and Natural Language Processing*, 37-102. Amsterdam/Philadelphia: John Benjamins.
- Napoli, Donna Jo. 1989. *Predication Theory: A Case Study for Indexing Theory*. Cambridge, U.K.: Cambridge University Press.
- Svensson, Patrik. 1989. *Number and Countability in English Nouns*. Uppsala: Swedish Science Press.
- Tesnière, Lucien. 1959. *Syntaxe Structurale*. Paris: Klincksieck.

Charles J. Fillmore  
International Computer Science Institute  
1947 Center St., Suite 600  
Berkeley, California 94704

fillmore@icsi.berkeley.edu

Hiroaki Sato  
Senshu University, Japan

hiroaki@isc.senshu-u.ac.jp